Using Machine Learning Techniques to Explore ^{1}H-MRS data of Brain Tumors

Félix Fernando González-Navarro and Lluís A. Belanche-Muñoz
Dept. de Llenguatges i Sistemes Informàtics
Universitat Politècnica de Catalunya
Barcelona, Spain
$\{fgonzalez, belanche\}$@*lsi.upc.edu*

***Abstract*—Machine learning is a powerful paradigm to analyze Proton Magnetic Resonance Spectroscopy (^{1}H-MRS) spectral data for the classification of brain tumor pathologies. An important characteristic of this task is the high dimensionality of the involved data sets. In this work we apply filter feature selection methods on three types of ^{1}H-MRS spectral data: long echo time, short echo time and an *ad hoc* combination of both. The experimental findings show that feature selection permits to drastically reduce the dimension, offering at the same time very attractive solutions both in terms of prediction accuracy and the ability to interpret the involved spectral frequencies. A linear dimensionality reduction technique that preserves the class discrimination capabilities is additionally used for visualization of the selected frequencies.**

***Keywords*-Machine Learning; Feature Selection; Proton Magnetic Resonance Spectroscopy; Classification; Visualization;**

I. Introduction

Proton Magnetic Resonance Spectroscopy or ^{1}H-MRS is a non-invasive technique that provides information about the biochemical profile (metabolites and lipids) of brain tissue; ^{1}H-MRS data has the appearance of a plot of peaks along the x-axis, with the peak position depending on the resonant frequency of the associated metabolite. As a result this technique provides a valuable tool in discovering or telling from possible disease processes [1] and serves as a support to other diagnostic methods like Magnetic Resonance Imaging (MRI) in clinical practice [2]. An example of a ^{1}H-MRS data set is depicted in Fig. I.There are several examples in the literature that use machine learning (ML) techniques in telling different classes of brain tumors. First attempts using ^{1}H-MRS data in assessing human brain tumors *in vivo* are back to [3]. It was found that spectra differ significantly from normal brain spectra and between tumors by detecting the presence/absence of different metabolites. Even though no ML analysis of spectra was done in establishing these differences, it was concluded that ^{1}H-MRS spectroscopy may help to differentiate tumors for diagnostic and therapeutic purposes, limiting the need for invasive and risky diagnostic procedures such as biopsies. ML techniques have been used in order to *automate* the classification task. Linear discriminant analysis (LDA) (e.g. [5], [6]) and artificial neural networks (e.g. [7], [8]) are commonly used methods. Later studies perform dimensionality reduction, either considering the *peak* signals, ratios between peak signals or *feature extraction* [14], [13]. First studies in performing explicit feature selection used a modification of the well-known Forward Selection method to select spectral points [9]. More recently, *filter* methods such as the Fisher criterion and the Relief algorithm, among others, have been considered [2].

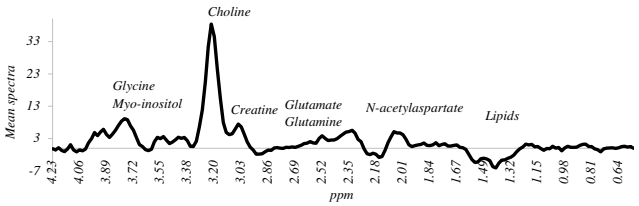

Figure 1. Example mean spectra for a long echo time ^{1}H-MRS data set. Some metabolites –resulting products of metabolic processes– are indicated in the spectra. For example: Creatine (tCre), a marker of oxidative metabolism of cells; Choline, which is a combination of multiple metabolites and is elevated in all brain tumors; Lipids (Lip), seen in condition of necrosis.

Due to the relatively high dimensionality of ^{1}H-MRS data (up to 500 spectral features), these efforts use dimensionality reduction methods such as PCA to lower the complexity of the problem; a recognized drawback is the difficulty to interpret the solution in original terms; moreover, PCA is not designed to preserve class separability [9]. However, data projection techniques that preserve the class discrimination achieved by a classifier are very welcome since they can be used to visualize the solutions generated by the feature selection process. In this work we develop a feature selection method to generate relevant subsets of spectral frequencies. The evaluation of subsets is based on entropic measures that treat single spectral frequencies as the features. The goal is to find attractive solutions in terms of low numbers of frequencies, high prediction accuracy and the interpretability of the selected spectral frequencies. The combination of feature selection and classification aims at obtaining simple models (in terms of low numbers of features) capable of good generalization. We also aim to progress in the comparison of MRS data acquired at different echo times, as well as in the comparison of these with data that combine both echo times.

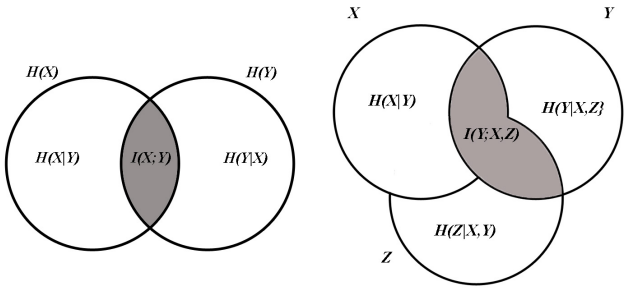

Figure 2. Left: Mutual information between two variables X, Y as shared conditional entropy. Right: Mutual information between two variables X, Z in the presence of a *class* variable Y.

II. Preliminaries

The *mutual information* (I) can be interpreted as a measure of the information that a random variable has or explains about another one:

$$I(X;Y) \quad = H(Y) - H(Y|X) = E_{X,Y}[log\, \tfrac{p(x,y)}{p(x)p(y)}] \quad (1)$$

where H denotes the *entropy*. Note that $I(X;X) = H(X)$, since $H(X|X) = 0$ and $I(X;Y) = I(Y;X)$. These concepts are graphically represented in Fig. 2 (left). It can be seen that by increasing $I(X;Y)$, $H(Y|X)$ decreases; in other words, there is a reduction in the uncertainty of Y due to the action of X. The computation of MI can be extended from the bivariate to the multivariate case, of a number $n \geq 2$ of variables against another one:

$$\begin{aligned} I(X_1, \ldots, X_n; Y) &= \sum_{i=1}^{n} I(X_i; Y | X_1, \ldots, X_{i-1}) \\ &= H(Y) - H(Y|X_1, \ldots, X_n) \quad (2) \end{aligned}$$

Conditional I is expressed in the natural way, by conditioning in (1):

$$I(X;Y|Z) = H(Y|Z) - H(Y|X,Z) \quad (3)$$

The MI in the bi-variate case has been successfully used in feature selection, as a way to measure the influence that a feature has over the class or target variable. Sometimes a normalized variant is used, given by $C_{XY} = \frac{I(X;Y)}{H(Y)}$, where Y is the class or target variable, commonly known as *coefficient of constraint* or *uncertainty coefficient* (note that the maximum value that $I(X;Y)$ can take is $H(Y)$). This coefficient has been used to create a measure of relevance for feature subsets with respect to a class feature (see e.g. [11]). Instead of conditioning, in this work we propose to calculate the MI between a class variable Y and two variables X and Z, as shown in Fig. 2 (right). The shaded area represents $I(Y; X, Z)$, the information that X, Z explain about Y:

$$\begin{aligned} I(Y; X, Z) &= H(Y) - H(Y|X,Z) \\ &= H(Y) + H(X,Z) - H(X,Y,Z) \quad (4) \end{aligned}$$

Given that $I(Y; X, Z) \leq 1$ and that $H(Y)$ acts as the baseline reference, it is wise to normalize it as:

$$R(Y; X, Z) = \frac{H(Y) + H(X,Z) - H(X,Y,Z)}{H(Y)} \quad (5)$$

An index of relevance R is obtained that evaluates the influence of two variables with respect to a class variable. It takes values between zero (no relevance) and one (maximum relevance). In order for this expression to be of practical use from the feature selection point of view, it needs to be extended to the multivariate case. The mutual information between a subset of variables and the class variable is computed by generating first a "super-feature", obtained considering the concatenation of each combination of possible values of its forming features. In symbols, let $X = \{X_1, ..., X_n\}$ be the full feature set and consider a subset $\tau = \{\tau_1, \cdots, \tau_k\} \subseteq X$. A single feature $\mathcal{V}_\tau$ can be obtained uniquely, whose possible values are the concatenations of all possible values of the features in τ (for completeness, define $\mathcal{V}_\emptyset = \emptyset$). The expression in eq. (5) can be determined as a simple *bivariate* case: an *index of relevance* of a feature $X_i \in X$ to a class Y with respect to a subset $\tau \subset X \setminus X_i$ is given by:

$$R(X_i; Y|\tau) = \frac{H(Y) + H(\tau, X_i) - H(\tau, Y, X_i)}{H(Y)} \quad (6)$$

The use of the super-feature allows faster computations, that are not essential for its understanding. A detailed implementation can be found in [19]. This way of calculating feature subset relevance is used to evaluate subsets of spectra, embedded into a filter *forward-search* strategy, conforming the BS^4 algorithm, a supervised filter independent of the search strategy or the *a posteriori* inducer, described below.

A. Algorithm BS^4: Best Spectral Subset Search Strategy

Algorithm BS^4–which stands for *Best Spectral Subset Search Strategy*–, is to be used with the index of relevance defined in (6); see **Algorithm 1**. It begins by selecting the best feature that maximizes its relevance with respect to the class feature (line 2-4). Then a forward search is conducted in which, at every step, the feature that provides the maximum value of relevance when added to the current subset is selected (line 7). It is possible that, at the end of a step, more that one feature renders the same value for relevance (lines 8-12). If this is the case, then the feature that produces the *minimum redundancy* of information is chosen (line 9). In case this newly added feature brings a positive benefit, it is added to the current subset (lines 15-16). The algorithm stops whenever the index of relevance

can no longer be augmented: either the maximum value has been reached, no increment in current relevance is possible or we have run out of features (line 20). This work was aimed to achieve results that reflect the true behavior of the system as much as possible; in other words, to obtain *reliably* relevant features. We therefore advocate for the use of *bootstrapping* techniques in the feature selection process. Bootstrap resampling techniques are used to yield mean performance estimates and their variability, and thus a more reliable measure of predictive ability. These methods are also well-suited for the construction of standard error estimates and confidence intervals when sample size is small or the distribution of the statistic is unknown.

Algorithm 1 $BS4$ Best Spectral Subset Search Strategy

1: **Input**: $X = \{X_1, \ldots, X_n\}$: Feature set;
C: Class feature
2: $\phi \leftarrow \underset{f \in X}{\operatorname{argmax}} \dfrac{H(C) - H(f, C)}{H(C)}$
3: $\Phi \leftarrow \{\phi\}$: Current best subset
4: $R \leftarrow \frac{H(C) - H(\Phi, C)}{H(C)}$: Current best relevance
5: $S \leftarrow false$
6: **repeat**
7: $\phi \leftarrow \left\{ \underset{f \in X \setminus \Phi}{\operatorname{argmax}} \dfrac{H(C) + H(\Phi, f) - H(\Phi, C, f)}{H(C)} \right\}$
8: **if** $|\phi| > 1$ **then**
9: $\phi^+ \leftarrow \underset{f \in \phi}{\operatorname{argmin}} I(\Phi, f)$
10: **else**
11: $\phi^+ \leftarrow \phi$
12: **end if**
13: $R^+ \leftarrow \frac{H(C) - H(\Phi \cup \{\phi^+\}, C)}{H(C)}$
14: **if** $R^+ > R$ **then**
15: $R \leftarrow R^+$
16: $\Phi \leftarrow \Phi \cup \{\phi^+\}$
17: **else**
18: $S \leftarrow true$
19: **end if**
20: **until** $R^+ = 1 \ \vee \ S \ \vee \ |\Phi| = n$
21: **Output**: Φ : Best Spectral Subset (BSS)

The original ^{1}H-MRS data sets S were used to generate $B = 5,000$ bootstrap samples $S_1, \ldots, S_B$ that play the role of *training sets* in the feature selection process: each relevance value calculated in the algorithm BS^4 (lines 4, 7 and 13) is the *average* across the B bootstrap samples, i.e., the *average behavior* of a feature is used to guide and stabilize the algorithm.

III. Experimental work

A. Data sets

MRS provides a detailed metabolic fingerprint of the tumor-affected tissue that varies according to the *echo time* of the acquisition and can be used to characterize these pathologies. The echo time is an influential parameter in ^{1}H-MRS spectra acquisition. In short-echo time spectra (typically 20-40 ms) some metabolites are better evaluated –e.g. lipids, myo-inositol, glutamine and glutamate–. However, there may be numerous overlapping resonances –e.g. glutamate/glutamine at 2.2 ppm and NAA at 2.01 ppm– which make the spectra difficult to interpret [6]. A long-echo time (270-288 ms) yields less metabolites but also less baseline distortion, resulting in a more readable spectrum. The analyzed ^{1}H-MRS data were extracted from a database resulting from the INTERPRET European research project. For details on data acquisition and processing, please refer to [12]. The basic data characteristics are detailed as follows:

- 195 single voxel *long-echo time LET* spectra acquired *in vivo* from brain tumor patientsy, including: meningiomas (55 cases), glioblastomas (78), metastases (31), astrocytomas Grade II (20), oligoastrocytomas Grade II (6) and oligodendrogliomas Grade II (5);
- 217 single voxel *short-echo time SET* spectra: meningiomas (58 cases), glioblastomas (86), metastases (38), astrocytomas Grade II (22), oligoastrocytomas Grade II (6) and oligodendrogliomas Grade II (7).
- 195 single voxel *long/short-echo time LSET* spectra, obtained by merging the 195 common observations of the two previous data sets.

Class labeling was performed according to the World Health Organization for diagnosing brain tumors by histopathological analysis of a biopsy sample. Both spectra were grouped into three superclasses: high-grade malignant tumors (metastases and glioblastomas), low-grade gliomas (astrocytomas, oligodendrogliomas and oligoastrocytomas) and meningiomas.

B. Experimental setup

In order to compute the necessary entropies, a discretization process is needed. This change of representation does not often result in a significant loss of accuracy (sometimes significantly improves it [17], [18]); it additionally offers reductions in learning time [20]. In this work, the data sets were discretized using the CAIM algorithm, which was selected for two reasons: it is designed to work with supervised data, and does not require the user to define a specific number of intervals [21]. First, algorithm BS^4 is applied to the discretized bootstrap samples to obtain the BSSs (one for each data set). Then the classifier development stage is conducted using the continuous frequencies (both in the full set and in the obtained BSSs). Five different classifiers were evaluated by means of 10 times 10-Fold Cross Validation (10x10cv): the nearest-neighbor technique with Euclidean metric (kNN) and parameter *k* (number of neighbors), the *Naïve Bayes classifier* (NB), a *Linear Discriminant classifier* (LDC), *Support Vector Machine with linear kernel* (lSVM) and parameter *C* (regularization constant) and *Support Vector Machine with radial basis function kernel* (rSVM) and parameter *C* (regularization constant) and σ^2 (smoothing in the kernel function).

Data set	$\|BSS\|$	CPU time	ppm
LET	16	5min	1.27, 1.23, 1.32, 1.21, 1.30, 1.17, 1.51, 1.55, 2.14 3.77, 3.03, 0.92, 1.44, 0.94, 2.20, 2.94
SET	16	5min	2.37, 1.32, 1.29, 2.35, 2.31, 1.34, 1.25, 1.23, 2.25 2.24, 1.38, 3.77, 3.81, 2.54, 3.03, 3.60
LSET	17	12min	S2.37, S1.27, S1.23, S1.30, S2.31, S2.39, S2.43 S1.36, L1.27, S3.05, S2.22, S3.03, S2.27, S3.55 S2.48, S3.77

Table I
FINAL SELECTED BSSS PER DATA SET. LEFT TO RIGHT READING INDICATES THE ORDER OF SELECTION IN THE FEATURE SELECTION PROCESS. FOR LSET DATA, THE ORIGIN IS INDICATED BY A PREFIX (L OR S).

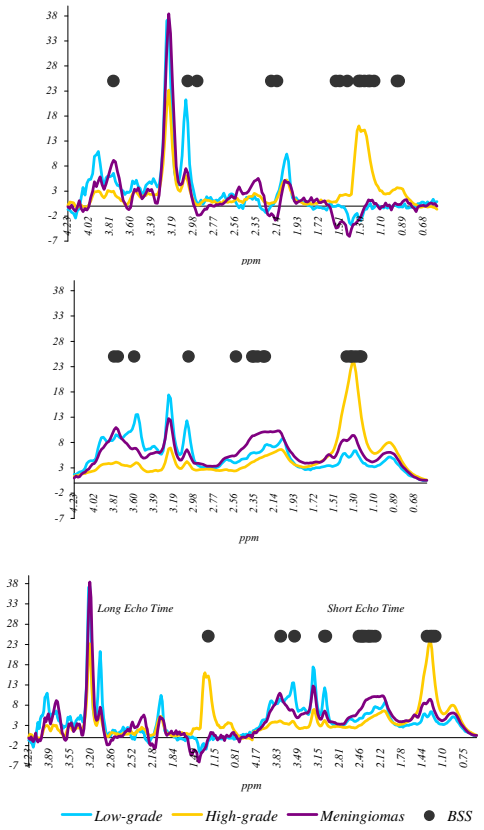

Figure 3. BSS for LET(Top), SET (Middle), LSET (Bottom) data set. Mean spectra for 3 classes are indicated.

C. Results and discussion

The results of the feature selection process are presented in Table I: for each ^{1}H-MRS data set, the obtained BSSs are detailed. Note the drastic reduction in the dimension and that subset sizes are very similar. For LSET data, all but one of the spectral points belong to the SET spectrum. Overall computation times are also shown. The positions of these selected spectral points in the spectrum are indicated in Figure 3. In LET, mostly fall in Glycine Myo-inositol region (3.77 ppm), Creatine peak (3.03, 2.94), Glutamete-Glutamine compunds (2.20, 2.14), Alanine peak (1.55, 1.51) and the rest fall in Lactate en Lipids regions. In SET,

Data set	NB	kNN	LDC	lSVM	rSVM
	Using BSS				
LET	83.2±0.06	86.0±0.06	83.0±0.08	87.9±0.09	**88.0±0.05**
SET	86.0±0.05	87.9±0.04	88.8±0.09	89.4±0.08	**89.6±0.06**
LSET	85.1±0.05	89.4±0.05	86.7±0.12	90.1±0.03	**91.0±0.05**
	Using the whole spectrum				
LET	84.9±0.09	86.3±0.05	79.8±0.22	89.9±0.04	**90.0±0.06**
SET	82.6±0.06	87.8±0.06	84.5±0.13	90.0±0.08	**90.4±0.08**
LSET	15.9±0.01	89.4±0.09	*	90.6±0.04	**90.7±0.08**

Data set	NB	kNN	LDC	lSVM	rSVM
	Using BSS				
LET	(82.7,83.6)	(85.5,86.5)	(82.4,83.7)	(87.2,88.6)	(87.7,88.5)
SET	(85.6,86.4)	(87.5,88.1)	(88.0,89.5)	(88.7,90.0)	(89.1,90.0)
LSET	(84.7,85.5)	(89.0,89.9)	(85.7,87.7)	(89.8,90.4)	(90.4,91.2)
	Using the whole spectrum				
LET	(84.1,85.6)	(85.9,86.7)	(78.0,81.6)	(89.5,90.2)	(89.5,90.5)
SET	(82.2,82.9)	(87.3,88.3)	(83.5,85.6)	(89.4,90.6)	(89.8,91.0)
LSET	(15.7,16.0)	(88.6,90.1)	n.a.	(90.3,90.9)	(90.0,91.3)

Table II
TOP:10X10CV ACCURACY (IN %) AND STANDARD ERROR PER DATA SET AND CLASSIFIER USING THE SELECTED SUBSETS OF SPECTRAL POINTS (TABLE I) AND THE WHOLE SPECTRUM. BOTTOM: CONFIDENCE INTERVALS. ASTERISK INDICATES NUMERICAL PROBLEMS.

True Class	LSET(17)+rSVM		
	LG	HG	ME
LG	7	1	0
HG	1	29	0
ME	0	1	14

True Class	LSET(390)+rSVM		
	LG	HG	ME
LG	7	1	0
HG	1	29	0
ME	0	1	14

Table III
TEST SET CONFUSION MATRIX FOR LSET WITH 17 FEATURES –SEE TABLE I– AND RSVM INDUCER; AND LSET USING THE WHOLE SPECTRUM (390 FEATURES) AND RSVM.

spectral points are in Glutamate-Glutamine and alanine region (3.81, 3.77), Myo-Inositol and Glycine (3.60), Creatine (3.03), Glutamate-Glutamine and N-acetylaspartate (2.54, 2.37, 2.35, 2.31, 2.25, 2.24) and the rest in the Lactate and Mobile Lipids region.

The 10x10cv accuracy rate is presented in Table II for each data set. It is seen that the rSVM apparently delivers the best results. Special attention deserves LSET where an accuracy as high as 91% is achieved, exceeding other accuracies reported elsewhere –see section D. The lSVM and kNN classifiers also show a good performance. A Wilcoxon Signed-Rank Test is performed for the (null) hypothesis that the median of the differences between rSVM errors and the error of another classifier is zero, for every data set –see Table IV. As it turns out, this hypothesis has to be rejected at the 95% level when rSVM is compared against all other classifiers, remarkably for all three data types. Also the p-values indicate that classification capacities using the whole spectrum of features and the rSVM classifier in the LSET data set are approximately the same –Table IV, right. This gives a strong support to the fact that by reduction of this problem to a very low dimensionality, a description of the phenomenon as accurate as using the entire set of can be achieved.

In a medical context, data visualization in a low-dimensional representation space may become extremely

Data set (Classifier)	NB	kNN	LDC	lSVM
LET (rSVM)	0.002	0.002	0.002	0.496
SET (rSVM)	0.002	0.002	0.049	0.922
LSET (rSVM)	0.002	0.004	n.a.	0.002

Data set (Classifier)	p-value
LET (rSVM)	0.002
SET (rSVM)	0.020
LSET (rSVM)	0.695

Table IV
WILCOXON SIGNED RANK TEST p-VALUES: LEFT: DIFFERENCES BETWEEN THE RSVM VS. THE REST IN THE SELECTED SUBSETS. RIGHT: DIFFERENCES BETWEEN THE RSVMS VS. THEMSELVES USING THE WHOLE SPECTRUM.

important, helping radiologists to gain insights into what undoubtedly is a complex domain. We use in this work a method based on the decomposition of the scatter matrix - arguably a neglected method for dimensionality reduction- with the remarkable property of maximizing the separation between the projections of compact groups of data (tumor classes, in this work). This is a recently introduced method that leads onto the definition of low-dimensional projective spaces with good separation between classes, even when the data covariance matrix is singular; further details about this method can be found in [22]. Such visualization is illustrated by the plots in Figure 4. These are scatter plots of 2-D projections of the three classes (using the first two eigenvectors of the scatter matrices).

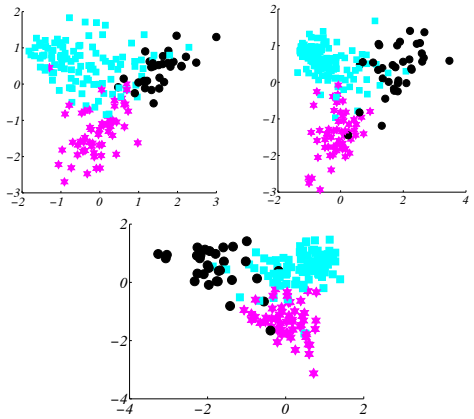

Figure 4. Projection of the data sets (using the *selected* feature subsets) onto the first two eigenvectors of the scatter matrices as coordinate system. Top-Left: LET, Top-right: SET, Bottom: LSET. Circles represent low-grade *gliomas*; filled squares high-grade *malignant tumors* and stars *meningiomas*.

D. Past usage of these data sets

Previous work by [14] using PCA followed by LDC *in LET data* showed that in distinguishing between high-grade malignant tumors and meningiomas, a mean area under the ROC curve of 0.94, using 6 principal components (PC) was reached. The same method was used to distinguish between high-grade malignant tumors and astrocytomas Grade II (part of the low-grade gliomas super-class), obtaining a mean AUC of 0.92, also using 6 PCs. [15] explored several feature extraction methods, including ICA, PCA and wavelet transformations and combined them systematically with different classifiers such as LDC, SVMs and ensemble methods; results offered a 91% classification accuracy using regularized linear regression, but in two-class problems. Regarding *SET data*, [16] used PCA (10 PCs) , with LDC, kNN and a Least-Squares SVM, reaching an 85% of accuracy in the same three super-class problem; [14] reported a mean AUC slightly than that for LET data: 0.95 with 4 PCs in the high-grade malignant tumors and meningiomas problem, and 0.97 with 3 PCs in high-grade malignant tumors and astrocytomas Grade II problem. In [10], PCA, Relief and a simple stepwise method were used alongside with LDC and the LS-SVM. Experimental results were: the combination LSET yielded 88.7% of accuracy; using LET data only, 82.5% of accuracy; using SET data only, 88.8%.

E. Summary and discussion of findings

In view of all these experimental results, several findings are now summarily presented:

1) Feature selection appears to be a viable avenue for dimensionality reduction. With about a tenth of spectral frequencies, the selected classifiers obtain mean performance figures close or better to those using the full set. This is specially important in the case of the rSVM technique.
2) As stated in the introduction, most of the existing literature indicates an advantage in using SET information over LET [5]. The present work adds strong support to this finding, given that all the classifiers obtained markedly better results for SET data against LET. The combined use of both echo times yields similar performance but using a lower number of frequencies.
3) The BS^4 algorithm seems to work well, as it produces rather sensible results with the reduced subsets. Its simplicity and fast performance, independently from the classifier or validation method, makes it a flexible solution in analyzing this specific domain.
4) The resulting sets of selected spectral frequencies (Fig. 3) happen to be located in truly biologically relevant regions, which permits an acceptable interpretation for the solutions.
5) The discrimination ability of these sets can also be subject of visual interpretation (Fig. 4), which definitely potentiates the supporting role to clinical practice.

IV. CONCLUSIONS AND FUTURE WORK

MRS is yet to become a standard method for day-to-day clinical diagnosis of brain tumors, despite being a non-invasive technique and one that provides rich information about the biochemistry of the tumor pathology. In this study, we report experimental results that support the practical advantage of combining robust feature selection and classification ML technics in this field of research. An attractive accurate classification is obtained with parsimonious and interpretable subsets of spectral frequencies. A linear dimensionality reduction technique that provides a data projection

–while preserving the class discrimination achieved by a classifier– is also used in our study. The developed proposal provides a drastic reduction in dimensionality while being competitive or even sometimes improving on the performance obtained using the full set of spectral frequencies. This holds true for all the acquisition modalities considered: short and long echo times, and the combination of both by concatenation of spectra. These results are extremely important as they make the diagnosis easily interpretable in terms of a handful of spectral frequencies, most of them associated to known metabolites. We have reached a step beyond feature selection in improving the interpretability of the results by providing a visualization method that preserves the discrimination capability of the obtained classifier models. Our classification results provide partial support to similar studies in the literature as they show the comparative advantage of using SET data or a combination of SET and LET data. Future research will extend the use of the proposed methodology to the analysis of other brain tumor classification problems involving different pathologies and pathology groupings.

ACKNOWLEDGMENTS

Authors gratefully acknowledge the former INTERPRET partners (INTERPRET, EU-IST-1999-10310) and, from 1st January 2003, Generalitat de Catalunya (grants CIRIT SGR2001-194, XT 2002-48 and XT 2004-51); data providers: Dr. C. Majós (IDI), Dr.A. Moreno-Torres (CDP), Dr. F.A. Howe and Prof. J. Griffiths (SGUL), Prof. A. Heerschap (RU), Dr. W. Gajewicz (MUL) and Dr. J. Calvar (FLENI); data curators: Dr. A.P. Candiota, Ms. T. Delgado, Ms. J. Martín, Mr. I. Olier, Mr. A. Pérez and Prof. Carles Arús (all from GABRMN-UAB). Authors also acknowledge funding for CICyT TIN2006-08114 and SAF2005-03650 projects; the Mexican CONACyT and Baja California University and thank the anonymous reviewers for their helpful suggestions.

REFERENCES

[1] Sibtain, N., et al.: The clinical value of proton magnetic resonance spectroscopy in adult brain tumors. *Clinical Radiology* **62**, 109–119 (2007)

[2] Luts, J., et al.: A combined MRI and MRSI based multiclass system from brain tumor recognition using LS-SVMs with class probabilities and feature selection. *Artificial Ingelligence in Medicine* **40**, 87–102 (2007)

[3] Bruhn, H., et al.: Noninvasive differentiation of tumors with use of localized h-1 MR spectroscopy in vivo: initial experience in patients with cerebral tumors. *Radiology* **172**, 541–548 (1989)

[4] Hagberg, G.: From magnetic resonance spectroscopy to classification of tumors. a review of pattern recognition methods. *NMR in Biomedicine* **11**, 148–156 (1998)

[5] Tate, A., et al.: Development of a decision support system for diagnosis and grading of brain tumors using in vivo magnetic resonance single voxel spectra. *NMR in Biomedicine* **19**, 411–434 (2006)

[6] Majos, C., et al.: Brain tumor classification by proton MR spectroscopy: Comparison of diagnostic accuracy at short and long TE. *American Journal of Neuroradiology* **25**, 1696–1704 (December 2004)

[7] Usenius, J., et al.: Automated classification of human brain tumors by neural network analysis using in vivo 1h magnetic resonance spectroscopic metabolite phenotypes. *Neuroreport* **7**(10), 1597–1600 (1996)

[8] Ala-Korpela, M., et al.: Artificial neural network analysis of 1h nuclear magnetic resonance spectroscopic data from human plasma. *Anticancer Research* **16**, 1473–1480 (1996)

[9] Nikulin, A., et al.: Near-optimal region selection for feature space reduction: novel preprocessing methods for classifying MR spectra. *NMR in Biomedicine* **11**, 209–216 (1998)

[10] Garcia, J., et al.: On the use of long TE and short TE SV MR. spectroscopy to improve the automatic brain tumor diagnosis. Technical report, In `ftp://ftp.esat.kuleuven.ac.be/pub/SISTA/ida/reports/07-55.pdf` (2007)

[11] Wang, H.: Towards a Unified Framework of Relevance. PhD thesis, U. of Ulster (1996). In `http://www.infl.ulst.ac.uk`.

[12] INTERPRET: International network for pattern recognition of tumors using magnetic resonance. `http://azizu.uab.es/INTERPRET`

[13] Lukas, L., et al.: Brain tumor classification based on long echo proton MRS signals. *Artificial Intelligence in Medicine*, 31, 73-89, (2004)

[14] Devos, A.: Quantification and classification of MRS data and applications to brain tumor recognition. PhD thesis, Katholieke Univ. Leuven, Belgium (2005)

[15] Menze, B., et al.: Optimal classification of long echo time in vivo magnetic resonance spectra in the detection of recurrent brain tumors. *NMR Biomedicine* **19**(5), 599–609 (2006)

[16] Ladroue, C.: Pattern Recognition Techniques for the Study of Magnetic Resonance Spectra of Brain tumors. PhD thesis, St. George's Hospital Medical School, United Kingdom (2003)

[17] Ng, M., Chan, L.: Informative gene discovery for cancer classification from microarray expression data. In: IEEE Workshop on Machine Learning for Signal Processing, IEEE, 393–398 (2005)

[18] Potamias, G., et al.: Gene selection via discretized gene-expression profiles and greedy feature-elimination. In: SETN., 256–266 (2004)

[19] F. González and Ll. Belanche. Gene subset selection in microarray data using entropic filtering for cancer classification. *Expert Systems*, 26(1), 113-124, 2009.

[20] Catlett, J.: On changing continuous attributes into ordered discrete attributes. In: Procs. of the European working session on learning on Machine learning, New York, NY, USA, Springer-Verlag New York, Inc., 164–178 (1991)

[21] Kurgan, L.A., Cios, K.J.: Caim discretization algorithm. *IEEE Trans. on Knowledge & Data Engineering* **16**(2), 145–153 (2004)

[22] Lisboa, P., et al.: Cluster based visualisation with scatter matrices. *Pattern Recognition Letters* **29**(13), 1814–1823 (2008)